Volume 6 • Number 1 • 2003
VALUE IN HEALTH

Making Decisions on Technology Availability in the British National Health Service—Why We Need Reliable Models

Ron L. Akehurst, PhD

School of Health and Related Research (ScHARR), The University of Sheffield, Sheffield, UK

Elsewhere in this issue, Dr. Milton Weinstein and colleagues set out principles for good practice in decision analytic modeling in health care evaluation [1]. The article is the most recent in a small series that have set out to assist both producers and users of models in their quest for excellence or at least adequacy. It builds on what has gone before and is the most comprehensive and valuable to date.

As a member of the Appraisal Committee of the National Institute for Clinical Excellence (NICE) in England, I am charged with fellow committee members to form views on whether particular health technologies (drugs, surgical procedures, diagnostic techniques, devices, educational programs, etc.) are effective and cost-effective and therefore whether they should be available, free of charge. When the committee was first formed, modeling was not regarded as having an important place in helping the committee reach decisions. Now, after three years' experience, its place is central, with the majority of appraisals being informed by models produced either by manufacturers or a NICE-commissioned assessment team.

Models play a number of important roles. They often enable estimation of long-term impacts of concern (mortality, quality of life) in any therapeutic area from the specific evidence of short-term efficacy available from clinical trials of the technology under consideration. Typical examples of the latter might be blood lipid levels, presence or absence of abnormal cells, or event rates, such as frequency of acute rejection episodes among solid organ transplant recipients or frequency of second myocardial infarctions in a group of previous sufferers. The modeling allows consideration of likely longer-term consequences of short-term effects as well as the combination of a variety of impacts (for example, from the various sequelae of diabetes).

For the committee, this facilitates comparison between different therapeutic areas by providing estimates of, for example, cost per quality-adjusted life year (QALY) gained as well as those, such as cost per acute rejection episode avoided, which might follow relatively readily directly from clinical trial evidence. The common currency of the cost per QALY is extremely helpful when technologies that are necessarily disparate are compared.

Even when data relevant to reaching a decision are scarce, modeling has proved useful. At times it has helped focus the committee on key parameters in a decision so that energy can be directed toward consideration of their likely scale. Thus, when it has been shown that the key issue in assessing cost-effectiveness is the quality-of-life impact of the clinical changes brought about, the committee has tried to assess whether threshold levels of gain are likely to be achieved or not. At times it has even proved possible to undertake supplementary collection of evidence on just the crucial issue.

More generally, models have proved their worth in indicating both the value of obtaining further information and what its nature should be. On at least one occasion, a formal value of information analysis included with a model has led to the committee recommending that a technology be supported for a period but in such a way that the indicated information was collected to inform a subsequent review of the decision.

Models have become an indispensable part of the analysis that the NICE Appraisal Committee uses to reach many of its decisions. It is therefore crucial that the models used are of high quality. For other types of evidence considered by the committee, there are in existence checklists of quality. Notably, the work of the Cochrane Collaboration and others has systematized the reviewing and synthesis of clinical trial evidence so that we can say not only what the evidence on a particular technology suggests is the case but also the strength of belief we should have in the result. We have yardsticks by which we can judge the quality of trials. This is only gradually becoming the case for models and the article by Weinstein and colleagues is a further step on the road to a generally held view of what constitutes good quality in modeling.

The Weinstein article represents the results of the deliberations of an ISPOR task force set up expressly for the purpose. The group consisted of

 1098-3015/03/$15.00/3 3–5

representatives from most of the main centers around the world engaged in this kind of activity on behalf of governments, industry, or both. Most of what has been learned about modeling for the purposes of health-care evaluation is embodied in the representatives of these centers.

The main part of the article consists of a statement of a group consensus regarding the attributes that define a good health-care decision model, under the headings of structure, data, and validation. The great merit of the article is that it is very extensive in the issues it considers—covering fine detail of the internal logic and arithmetic at one extreme to processes for justifying the use of the particular data chosen at another. The section on validation is a helpful development of what has gone before and is particularly important to groups such as NICE. There is a comfort in knowing that we understand why model *x* gives different predictions to model *y* because it reassures us that we have grasped key issues in appraising a technology. Equally, there is comfort in knowing that predictions from a model have been tested as far as possible and it has not been found wanting. As a decision maker, the fear is that the model relied on floats in a world of its own with no test in reality.

Areas for Yet Further Development

Excellent though it is, the article does not provide a cookbook for assessing models. It provides a very useful list of questions to ask and how to ask them, but an analyst is clearly left to judge how much less than perfection in one of the areas highlighted matters to the utility of a particular model. In the same way an analyst must judge whether a weakness in design of a clinical trial is of consequence in a particular context—for example, how limitations in inclusion criteria affect the transferability of conclusions on efficacy to a wider population—judgments must be made about models. When is a prediction made by a model close enough to observable data? This is the art of evaluation and decision taking and cannot be captured in a short article, if at all. Nevertheless, the hints that judgments of this type must be made that surface obliquely in the Weinstein article are welcome and deserve to be reinforced.

From the point of view of a user of models on the NICE Appraisal Committee, there are three particular directions in which commentary could usefully be strengthened. First, in common with all economic evaluations, the specification of the question is crucial, as is making it clear to users of the model what that question is. Thus, a model may give some clear messages about the potential cost-effectiveness of a technology, but the (implicit) question being addressed is not relevant to any policy or clinical decisions being contemplated. This suggests that the process of determining exactly what question should be addressed is itself something which could and should be made systematic and transparent. Although Weinstein and colleagues in their discussion of model structures touch on it, this part of process, which is crucial to decision making, is deserving of further attention. A section on the exact question being addressed, and why, is a must in reports which accompany models.

Second, and related, the comparators that are included in the model really matter. Again, the discussion of the process by which comparators should be chosen is fairly slight, but if comparators are included that are not relevant in a particular decision context, the model may be worthless. It is easy to pass off the choice of comparator as an issue of data availability. It is not; it is an issue of the decision to be informed. Differences between clinical practices in different countries (and even different hospitals in the same countries) may render models that are valuable in one national context useless in another. The question has arisen at times in the NICE appraisal context of the relevance of models developed outside the United Kingdom. There may be many difficulties in using models developed overseas. Few are insuperable but inclusion only of comparators which are not relevant in the United Kingdom is one of them.

Third, the issue of transparency is key. Weinstein and colleagues consider this but only briefly. During the NICE appraisal process there is a real reluctance to trust a model if it cannot be examined in detail, even if it has already been published and subject to peer review. Not all journals have a reliable peer review process for models and not all require the availability of the full model in electronic form. At the heart of this mistrust is a concern that parties to the NICE process might deliberately or inadvertently mislead the committee on cost-effectiveness estimates through misspecification of the model in some way.

There are, of course, understandable reasons why models might be withheld, for example, because they contain data that are commercial in confidence. This is certainly an issue for NICE. However, equally important is the interest of academics in intellectual property rights in a model. This is touched on only briefly by Weinstein and col-

leagues, but must be tackled or models will continue to be published with readers unable to check in detail what has been done.

A major part of the answer to the transparency problem may lie with the editorial policies of academic journals. If journals insist on true transparency and availability of models for examination before modeling papers are published, this may force the issue. A problem that journals have had until recently has been the absence of standards against which submitted models may be judged. Weinstein and colleagues have taken on those standards a further stage from those published before and give editors and decision makers alike a valuable tool in assessing the merits of any particular model. The article is greatly to be welcomed.

Reference

1 Weinstein MC, O'Brien B, Hornberger J, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices—Modeling Studies. Value Health 2003;6:9–17.